Imprecise and Qualitative Probability in Systems Biology

M. Sriram Iyengar, PhD[1*], Mary F. McGuire, MSc[1]
[1] School of Health Information Sciences, University of Texas Health Science Center at Houston
*email: m.sriram.iyengar@uth.tmc.edu

Introduction

Predictive modeling in biology is difficult because a) *randomness* is ubiquitous, and b) biological processes are profoundly complex, containing hundreds or thousands of component interactions. The latter leads to *uncertainty*, i.e., precise information about probabilities, pathway structure, rate constants and similar parameters, is often unknown. Yet, for many reasons, including drug development, it is important to accurately and objectively predict the course of biological processes under various conditions. Probabilistic models including Bayesian networks (Lucas, 2004) and stochastic pi-calculus (Priami *et al*, 2001) have been proposed in this regard.

However, since uncertainty differs qualitatively from randomness, and arises from different causes, probabilistic methods should not be expected to also model uncertainty. In fact, serious concerns have been identified about interpretation of models in which classical probability has been used to describe both uncertainty and randomness (Morgan and Henrion, 1990). Further, it is often impossible to assign precise point probabilities to each of the myriad constituents of an intricate biological pathway.

Here, we review imprecise probability and qualitative probability, techniques for predictive modeling that hold promise to address these issues.

Imprecise Probability

Imprecise Probability (ImP) offers intriguing potential to enhance systems biology with tools for rigorous predictive reasoning. ImP includes a spectrum of mathematical formulations. Some may be viewed as methods for sensitivity analysis of unknown or unknowable probabilities. Others comprise quantitative methods for reasoning with ambiguous or incomplete information.

A basic notion of ImP is *Interval Probability,* where point probabilities are replaced by intervals, e.g., 0.25 by [0.22, 0.28]. Formalizations and extensions of interval probability include *Random sets*, *Upper/Lower probability functions* and *Clouds* (Destercke *et al*, 2007). *Probability Bounds Analysis* (*P-boxes*) constitute an interval-type formulation of cumulative distribution functions (Ferson *et al*, 2003).

Other specifications include *Fuzzy reasoning*, *Plausibility functions*, *Possibility Functions*, and *Dempster-Shafer Theory,* a calculus addressing numerical degrees of belief and evidence. (DeCooman *et al*, 2007).

For predictive modeling of biological pathways, an interesting ImP formulation of Bayesian networks is *Credal networks,* based on *Credal sets,* a closed convex set of probability mass functions (Cozman, 2000). The theory includes definitions of (imprecise) conditional

probabilities and Bayes-type formulae for imprecise posterior probabilities. Similar to Bayesian networks, Credal networks are limited to direct acyclic graphs.

A Credal network representation of a signaling pathway (TNF-R) is being developed by the authors and appears capable of yielding meaningful predictions. We are also investigating techniques to overcome the acyclicity limitation.

Qualitative Probability

Qualitative probability(QP) may be appropriate when numerical approximations cannot be justified. QP replaces numerical values of likelihood by an ordinal set, i.e., an order relation on a set algebra (Luce and Narens, 1978). Wellman's specification of *qualitative probabilistic networks* (Wellman, 1990) includes a QP calculus for acyclic graphs, and *qualitative influences/synergies* analogous to conditional probability. Kappa calculus uses a five-point scale to assign qualitative probabilities (Darwiche and Goldszmidt, 1994).

Conclusion

Systems biology methods based on Imprecise Probability or Qualitative Probability can lead to useful predictive models robust with respect to the uncertainty and randomness typically encountered in biomedical research data.

References

Cozman FG (2000) Credal Networks. Artificial Intelligence 120:199-233

Darwiche A, Goldszmidt M (1994) On the relation between kappa calculus and probabilistic reasoning. Proc. Tenth Conf. on Uncertainty in Artificial Intelligence (UAI-94). 145–153

DeCooman G, Vejnarova J, Zaffalon M (eds) (2007) Proceedings of the Fifth International Symposium on Imprecise Probability, 155:164. Also see www.isipta.org.

Destercke S, Dubois D, Chojnacki E (2007) Relating practical representations to Imprecise Probabilities. Proceedings of the Fifth International Symposium on Imprecise Probability, 155:164. Also see www.isipta.org.

Ferson S, Ginzburg L, Kreinovich V, Myers D (2003) Construction Probability Boxes and Dempster-Shafer Structures Technical Report, Sandia National Laboratories

Lucas P (2004) Bayesian Analysis, Pattern Analysis, and Data Mining in Health Care. Current Opinion in Critical Care, 10:399-403

Luce RD, Narens L (1978) Qualitative Independence in Probability Theory. Theory and Decision 9: 225-239

Morgan, MG, Henrion M (1990) Uncertainty: A Guide to Dealing with Uncertainty in Quantitative Risk and Policy Analysis. Cambridge University Press, New York

Priami C, Shapiro CE, Regev A, Silverman W (2001) Application of a Stochastic Name-Passing Calculus to Representation and Simulation of Molecular Processes. Information Processing Letters, 80: 25-31

Wellman MP (1990) Fundamental Concepts of Qualitative Probabilistic Networks. Artificial Intelligence. 1990, 44 (3) 257 – 303